Multilevel Approach for Modeling TCP/IP

Urtzi Ayesta[*,†,*], Konstantin E. Avrachenkov[†,†], Eitan Altman[†], Chadi Barakat[†]. Parijat Dube[‡‡]

Abstract

In order to compute performance of both short and persistent TCP connections, both session level aspects as well as packet level aspects have to be taken into consideration. To that aim, we analyze with NS simulations the aggregated packet arrival process into a bottleneck queue generated by multiplexed TCP flows. We explain qualitatively the shape of the packet interarrival time distribution. In particular, we provide conditions under which the distribution of the inter packet arrivals is close to exponential and show how this condition scales when the network capacity becomes large. In addition, we analyze the structure of the autocorrelation function of times between packet arrivals. We point out the limitations of the NS simulator in the context of HighSpeed links. For the case of a packet arrival process close to Poisson, we develop a Fixed Point based model that allows us to compute the packet loss probability and the utilization of the bottleneck link.

Key words. Multiplexed TCP flows, HighSpeed links, Packet interarrival time distribution, Fixed Point Approach, NS simulations.

1 Introduction

We analyze with NS simulations [15] the packet arrival process into a bottleneck link generated by multiplexed TCP flows. Many different scenarios and parameter settings are considered: Different file size distributions, different loads, combination of short and persistent TCP connections, different link speeds, different number of access links, etc. We explain the structure of the packet interarrival time distribution. In particular we observe that there always exists a sufficiently small access link capacity that makes the interarrival time distribution to be close to exponential. We show that this property scales well when the capacity of the network increases. Still we observe that the packet interarrival times are not independent. We explain the location of the maximum of the autocorrelation function of these times.

For mathematical tractability, it is very important to identify under which conditions the aggregated TCP traffic is similar to the Poisson stream at the packet level. The traffic is closer to

[**]France Telecom R&D, 905 rue A. Einstein, 06921 Sophia Antipolis Cedex, France. e-mail: Urtzi.Ayesta@francetelecom.com

[††]INRIA Sophia Antipolis, 2004 route des Lucioles, B.P.93, 06902, Sophia Antipolis Cedex, France. e-mail: {uayesta, k.avrachenkov, altman,cbarakat}@sophia.inria.fr

[‡‡]IBM T.J. Watson Research Center, Yorktown Heights, NY 10598, USA

Poisson if we increase the number of access links that carry the TCP traffic to the bottleneck link. The traffic is also closer to Poisson if we slow the access links. For a given number of access links and a given speed of the bottleneck, we provide an empirical expression for the speed of the access links that leads to an aggregated packet arrival process at the bottleneck very close to Poisson. An interesting observation we make is that this expression is independent of the load ! The observation that the traffic at the packet level can be close to Poisson has been observed in another context, that of a network running at high load [7]. It has been shown in [7] that when the load of the network increases and due to multiplexing, the times between packet arrivals tend to be independent and their distribution tend to exponential. Our work deals more with the impact of network topology, especially the access links, on the packet arrival process in the core of the network.

In the case we are close to Poisson, we propose a Fixed Point Approach (FPA) which is applicable to the scenario where both persistent and short TCP connections are present in the network. The main idea behind FPA for TCP/IP networks is to combine a model for the IP network at the packet level with a model for the TCP connection performance based on some given packet loss process [4, 6, 10, 11]. By using FPA one can calculate the packet loss probability, the utilization of the bottleneck link and the throughputs of the persistent connections with good accuracy.

In the general case (even when the packet arrival process is not close to a Poisson process) we indicate bounds on the packet loss probability. The $M/M/1/K$ queuing model provides the lower bound and a queuing model corresponding to batch arrivals provides an upper bound. (The size of a batch is obtained by an aggregation of all packets from a source belonging to the same congestion window.)

Carrying out simulations with the NS simulator [15] for high speed links, we have observed a limitation of the current NS implementations. Namely, in NS, events are scheduled with a microsecond granularity. This may pose the problem of scalability for the simulations of high speed links where events needs to be scheduled at finer granularity than the microsecond.

The rest of the paper is organized as follows: In Section 2 the benchmark network model is introduced. In Section 3 the multiplexing of TCP flows is analyzed using NS simulations. In particular the limitations of NS for the modeling of high speed links is discussed. In Section 4 the FPA model and bounds on the packet loss probabilities are presented. The paper ends with a discussion and future research directions' section.

2 Benchmark network model

We study the aggregated packet arrival process and applicability of the FPA on the benchmark example of TCP/IP network with a single bottleneck (see Figure 1). This topology may for instance represent an access network. The capacity of the bottleneck link is denoted by C and its propagation delay is denoted by d. The capacities of N links leading to the bottleneck link are supposed to be large enough (or the load on each access link is small enough) so that they do not hinder the traffic. Each of these N links has a propagation delay d_i (the difference in propagation delays also

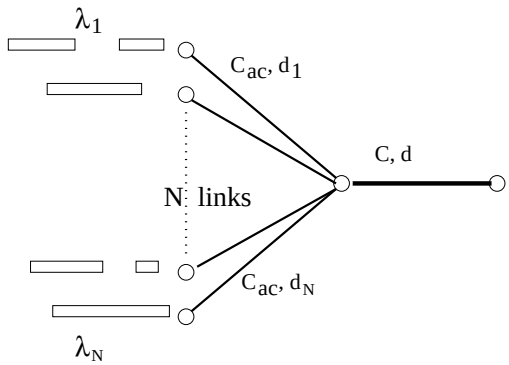

Figure 1: IP network with a single bottleneck link

improves multiplexing) and we assume that new TCP connections arrive on link i according to a Poisson process with rate λ_i. Thus, the *nominal* load of short TCP connections can be calculated as follows:

$$\rho_0 = \frac{E[doc.size] \sum_{i=1}^{N} \lambda_i}{C}, \tag{1}$$

where $E[doc.size]$ is the average document size to be transfered. We use the exponential and Pareto distributions for the document size (with $E[doc.size] = 10Kbytes$ and Pareto with infinite variance) [5, 14]. We also consider a scenario with the mixture of short and persistent TCP connections.

In the NS simulations we use the following values for the network parameters: bottleneck capacity – {1000,100,50,10,5,1}Mbps, bottleneck buffer size – {50,200} packets, bottleneck link propagation delay – 40ms, the number of access links – {10,50,100}, access link capacity is the same for all accesses links and is varied between 200Kbps and 2000Mbps , propagation delays of access links are uniformly distributed between 20 and 60ms, and the maximum segment size (MSS) is 500bytes. As for the buffer management, we consider Drop-Tail policy. It is still the most commonly used buffer management policy in the Internet. The buffer sizes of the access links are chosen large enough so that losses occur only in the bottleneck queue.

3 Simulation study of the multiplexed TCP flows

Let us study the input process at the bottleneck queue. In particular, we are interested under which conditions the aggregated traffic arriving to the bottleneck queue is close to Poisson. In the first subsection we consider a scenario with only short TCP flows. In the second subsection we consider a scenario with only persistent connections and another scenario with the mixture of short and persistent connections. Finally, in the third subsection we analyze the simulation of multiplexed TCP flows in the case of high speed links.

3.1 Short TCP flows

We choose the nominal system loads $\rho_0 = \{0.3, 0.6, 0.9\}$ (see equation 1). In Figures 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 we plot the distribution of the packet interarrival time for different parameter settings. In general, we observe that when the capacity of the access links is much smaller than the one of the bottleneck, the interarrival time distribution practically coincides with the exponential. For example, one can see that with 50 access links (Figures 3, 6), a capacity of 2Mbps and 200Kbps respectively make the packet interarrival time to be very close to exponential. Comparing Figures 2, 5 versus Figures 4, 7 we observe that the distribution of the packet interarrival times is closer to exponential when the number of access links increases. In Figures 8, 9, 10 we consider Pareto file size distribution for different nominal loads and in Figure 11 we consider Pareto with a larger average file size ($30KBytes$). All the above observations still hold. Thus, we conclude that with enough multiplexing, there exists such a "small enough" capacity for the access link so that the packet interarrival time distribution becomes very close to exponential. Namely, we noticed that, if the ratio NC_{ac}/C it is not large (around 2 in these particular settings) the distribution of the packet interarrival times is very close to exponential.

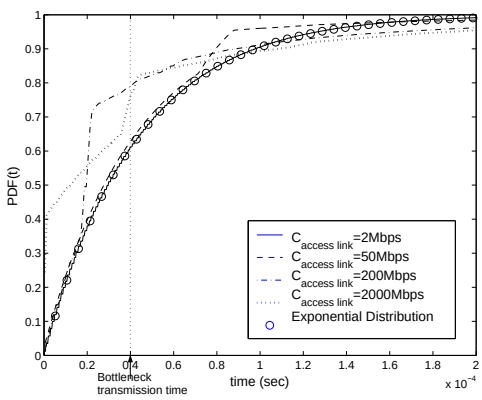

Figure 2: Bottleneck 100Mbps, 100 access links, Exponential file size and load 0.9.

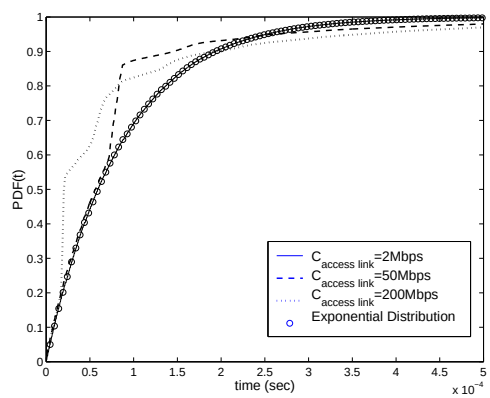

Figure 3: Bottleneck 50Mbps, 50 access links, Exponential file size and load 0.9

We would like to note that even in the case of slow access links the packet interarrival times are correlated (see Figure 12). In particular, we observe that the number of lags corresponding to the maximum value of the correlation is equal to the ratio between the access link transmission time and the average packet interarrival time at the bottleneck node. The interpretation for this is that the correlation is introduced by packet pairs sent in the Slow-Start phase from the same access link.

Let us now explain in more detail the case depicted in Figure 2. The same thinking applies to the other figures as well. In the case of slow access links ($C_{ac} = 2Mbps$) the interarrival time distribution practically coincides with the exponential. With the increase of the access link capacity the packet interarrival distribution starts to deviate from the exponential one. In particular, one can see the appearance of steps in the distribution function. In the case of access link capacities

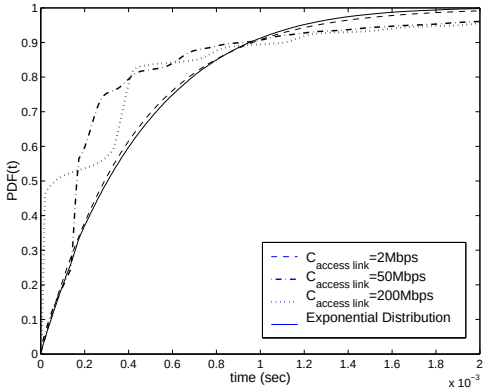

Figure 4: Bottleneck 10Mbps, 10 access links, Exponential file size and load 0.9

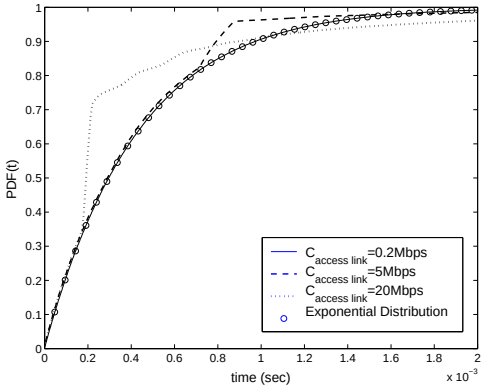

Figure 5: Bottleneck 10Mbps, 100 access links, Exponential file size and load 0.9.

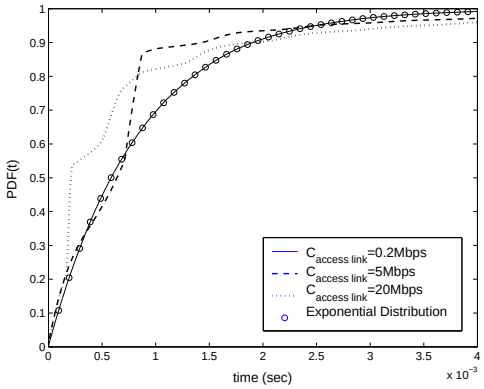

Figure 6: Bottleneck 5Mbps, 50 access links, Exponential file size and load 0.9

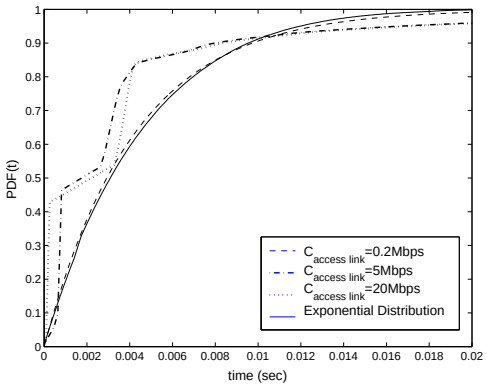

Figure 7: Bottleneck 1Mbps, 10 access links, Exponential file size and load 0.9.

smaller than the one in the bottleneck (see $C_{ac} = 50Mbps$) there is only one step corresponding to the transmission time of the access link. We can explain this again by the fact that packets come in pairs during the slow-start phase. When the access link capacity is greater than the bottleneck link capacity (see $C_{ac} = 200Mbps$ and $2000Mbps$) there are two steps. The first one corresponds to the transmission time of the access link and the second one corresponds to the transmission time of the bottleneck link. This second step can be interpreted as a typical time interval between two pairs of packets coming from the same access link. The above observations prompt us to approximate in the case of high access link the input process at the bottleneck queue as a batch arrival process (see Section 4 for more details).

3.2 Persistent and Short Connections

First let us analyze the multiplexing of only persistent connections. In Figure 13 we consider a bottleneck of 10Mbps, 10 access links and one TCP persistent connection on every access link. In Figure 14 we consider 100 access links also with one TCP persistent connection per each access link. We would like to note that there is a clear difference between the cases of only short and only

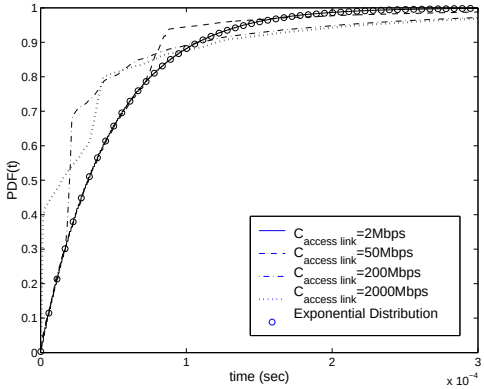

Figure 8: Bottleneck 100Mbps, 100 access links, Pareto file size and load 0.9

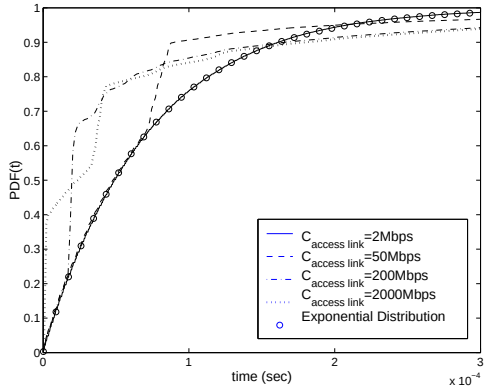

Figure 9: Bottleneck 100Mbps, 100 access links, Pareto file size and load 0.6

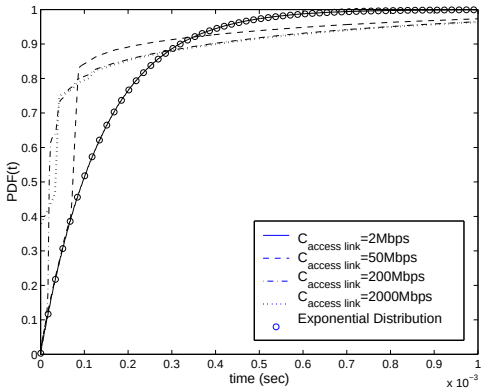

Figure 10: Bottleneck 100Mbps, 100 access links, Pareto file size and load 0.3

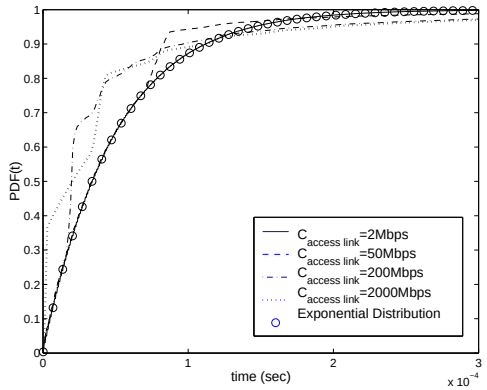

Figure 11: Bottleneck 100Mbps, 100 access links, Pareto file size, load 0.9, average file size 30KBytes

persistent connections. In the present case we observe that for large enough number of access links the interarrival time distribution is always relatively close to the exponential regardless the value of the capacity of the access link.

Despite the closeness of the interarrival time distribution to exponential, it is worthwhile to mention that as long as the capacity of the access link increases, so the loss probability does. In our particular setting (Figure 14), the loss probability is three time as larger for $C_{ac} = 50Mbps$,$200Mbps$,$2Gbps$ than for $C_{ac} = 2Mbps$. This can be explained by stronger correlation in the three former cases than in the later (see Figure 15).

Next we add to the above scenario short pareto distributed TCP connections. We consider two different settings. In the first one (Figure 16) the nominal load for short TCP connections is 0.8 and in the second one it is of 0.2 (Figure 17). Comparing Figures 16 and 17 with Figure 8, one can see that persistent connections smooth out the burstyness of the slow-start phase of short TCP flows.

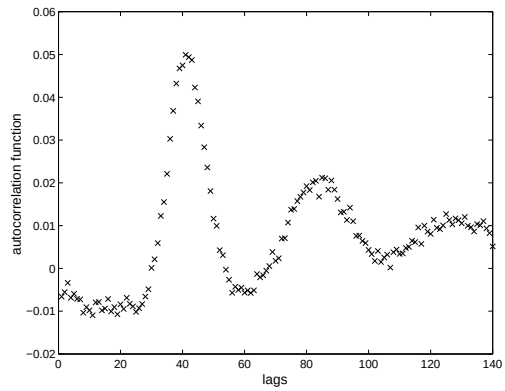

Figure 12: Correlation function of the packet arrival process:Bottleneck 100Mbps, 100 access links and $C_{ac} = 2Mbps$

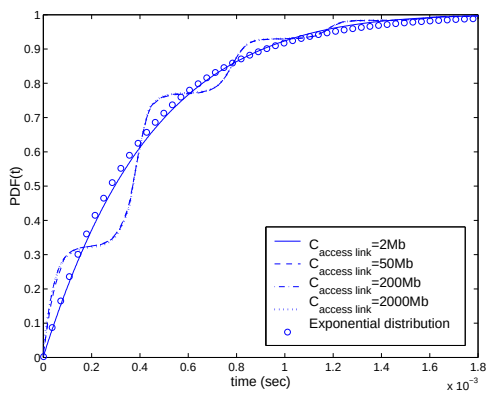

Figure 13: Bottleneck 10Mbps, 10 access links, Persistent connections

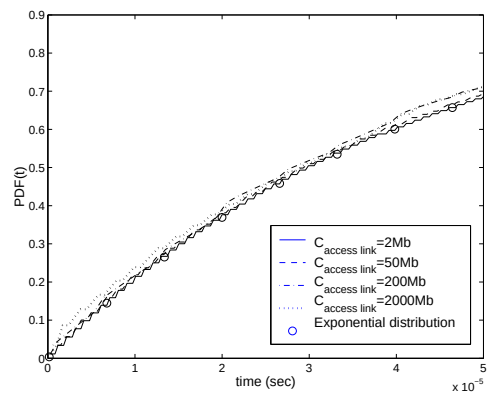

Figure 14: Bottleneck 100Mbps, 100 access links, Persistent connections

3.3 High-Speed Links

In this subsection we point out the limitation of NS simulator in the case of high-speed links. In NS, events are scheduled with a time granularity of microseconds. As a consequence, events that on continuous time are separate by less than one microsecond occur in NS at exactly the same time. We already noticed this effect in the case of 100Mbps bottleneck link (see in Figure 18). This effect becomes more pronounced when the link speeds increase. For example we plot in Figures 19 and 20 the simulation results for 1Gbps bottleneck with 100 and 10 access links respectively.

4 Fixed Point Approach

For the case of small ratio NC_{ac}/C we develop a Fixed Point Approach to compute the packet loss probability and the network utilization.

We recall that TCP is a protocol for reliable data transfer. Lost packets are retransmitted by the TCP until they are well received by the destination. Taking into account retransmissions, the

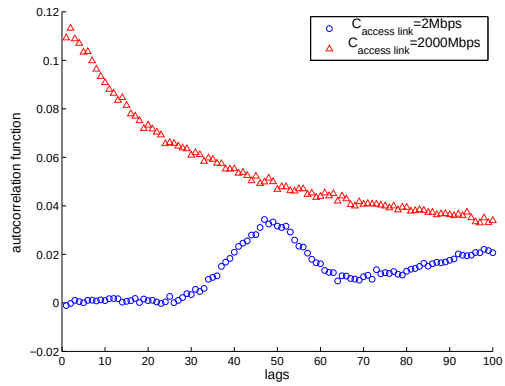

Figure 15: Autocorrelation function for the same scenario as in Figure 14

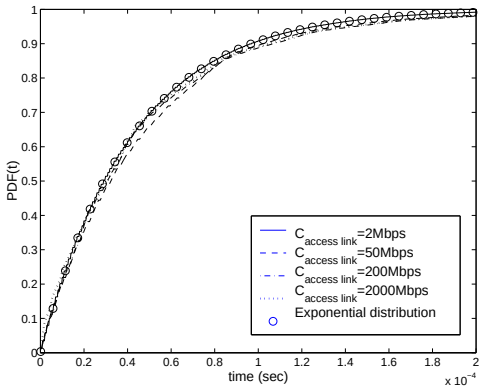

Figure 16: Bottleneck 100Mbps, 100 access links, Persistent and short connections. 80% of load due to short connections

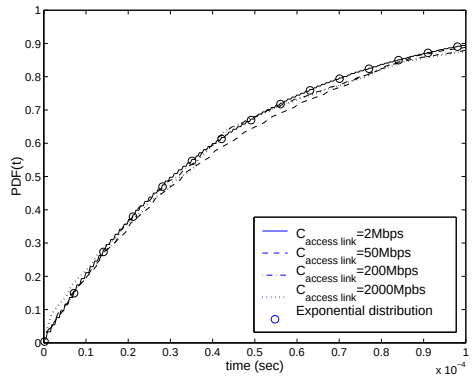

Figure 17: Bottleneck 100Mbps, 100 access links, Persistent and short connections. 80% of load due to short connections

actual load on the bottleneck link caused by short TCP flows is given by the formula [2]

$$\rho = \frac{\rho_0}{1-p},$$

where ρ_0 is the nominal load corresponding to the short TCP flows. In order to take into account the effect of persistent connections, the load can be modified as

$$\rho = \frac{\rho_0}{1-p} + \frac{M}{C}T(p), \tag{2}$$

where $T(p)$ is some expression for the average sending rate of a persistent TCP connection [1, 12, 16] and M is the number of persistent connections. In particular, one can use a new expression for the high speed TCP sending rate [9], whose asymptotic might be different from the square root formula that models the sending rate of current TCP implementations.

As for the buffer management, we consider the DropTail policy. This latter policy is still the most commonly used buffer management policy in the Internet. Our model can be easily extended

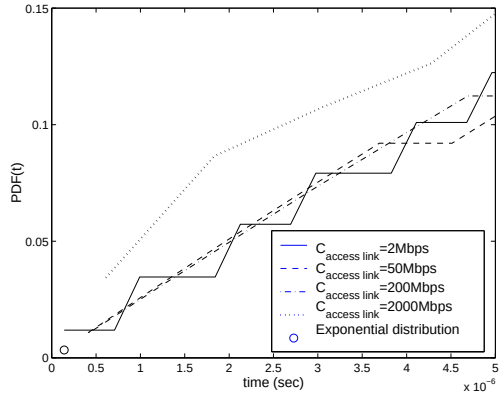

Figure 18: Zoom in of figure 16

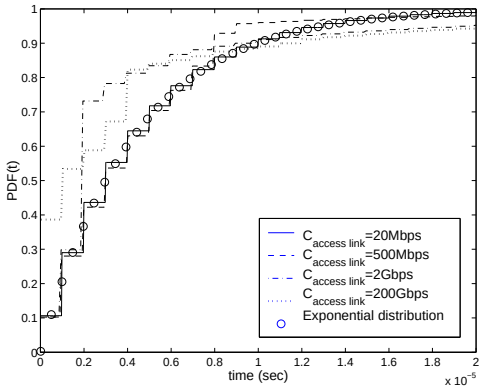

Figure 19: Bottleneck 1Gbps, 100 access links, Pareto file size, load 0.9, average file size 10KBytes

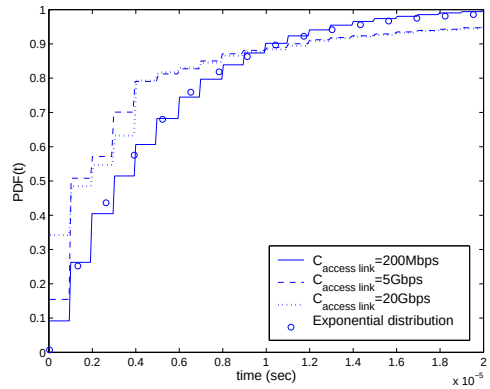

Figure 20: Bottleneck 1Gbps, 10 access links, Pareto file size, load 0.9, average file size 10KBytes

to other policies as RED. Assuming that the packets of the aggregated TCP traffic arrive to the bottleneck queue according to a Poisson process and the service time is exponentially distributed, one can use the classical $M/M/1/K$ queuing model. We take the assumption about the exponential service in order that our results will be analytically tractable. Thus, we have

$$p = \frac{\rho^K(1-\rho)}{(1-\rho^{K+1})}, \tag{3}$$

where K is the buffer size in packets.

Proposition 1 *If $\rho_0 < 1$ and $T(p)$ is a continuous non-increasing function of p, the system of equations (2) and (3) has a unique solution.*

PROOF: We substitute the expression for the packet loss probability (3) into (2) to get

$$\rho = \frac{\rho_0(1-\rho^{K+1})}{1-\rho^K} + \frac{M}{C}T\left(\frac{\rho^K(1-\rho)}{1-\rho^{K+1}}\right)$$

Multiplying by $(1-\rho^K)/(1-\rho^{K+1})$, we get

$$\frac{\rho(1-\rho^K)}{1-\rho^{K+1}} = \rho_0 + \frac{1-\rho^K}{1-\rho^{K+1}}\frac{M}{C}T\left(\frac{\rho^K(1-\rho)}{1-\rho^{K+1}}\right)$$

or, equivalently,

$$1 - \frac{1}{1+\rho+\rho 2+...+\rho^K} - \frac{1-\rho^K}{1-\rho^{K+1}}\frac{M}{C}T\left(\frac{\rho^K(1-\rho)}{1-\rho^{K+1}}\right) = \rho_0. \tag{4}$$

Let us consider the third term on the left hand side of the equation (4). The function $T(\rho^K(1-\rho)/(1-\rho^{K+1}))$ is non-increasing in ρ, since it is a composition of the increasing function $\rho^K(1-\rho)/(1-\rho^{K+1})$ and non-increasing function $T(p)$. Next, we note that $(1-\rho^K)/(1-\rho^{K+1})$ is a decreasing function in ρ. To show this, we prove by induction that the numerator of the derivative is less or equal than zero: $-K+\rho(K+1)-\rho^{(K+1)} \leq 0$. The induction step goes as follows.

$$-(K+1)+\rho(K+2)-\rho^{(k+2)} = -(K+1)+\rho(K+2)-\rho\rho^{(K+1)} \leq$$

$$\leq -(K+1)+\rho(K+2)-\rho(-K+\rho(K+1)) = -(K+1)(1-\rho)2 \leq 0.$$

Hence, as a product of the positive decreasing and positive non-increasing functions, the function

$$\frac{1-\rho^K}{1-\rho^{K+1}}\frac{M}{C}T\left(\frac{\rho^K(1-\rho)}{1-\rho^{K+1}}\right)$$

is positive non-increasing. Furthermore, since $(1-\rho^K)/(1-\rho^{K+1}) \sim 1/\rho$ as $\rho \to \infty$,

$$\frac{1-\rho^K}{1-\rho^{K+1}}\frac{M}{C}T\left(\frac{\rho^K(1-\rho)}{1-\rho^{K+1}}\right) \to 0, \quad \text{as} \quad \rho \to \infty.$$

The second term on the left hand side of the equation (4) also goes monotonously to zero as ρ goes to infinity.

Thus, we conclude that the left hand side of (4) is a non-decreasing function for $\rho \in (0,\infty)$ with a horizontal asymptote $y=1$. Hence, if $\rho_0 \geq 1$, there is no solution, and if $\rho_0 < 1$ there is a unique solution.

□

The system of nonlinear equations (2) and (3) could be solved by any standard numerical method.

For the general case, when the aggregated traffic can not be modeled by the Poisson process, we propose two bounds for the packet loss probability. In the most extreme case all packets from the same round can be considered as a single batch. Thus in this extreme case the distribution of the batch size is given by the distribution of the congestion window size. We compute this distribution assuming the session does not experience any loss and that it will remain always

in Slow-Start phase. This will give us an upper bound on the batch size. Conditioning on the number of rounds the probability of having a window of size $w_i \in \{1, 2, 4, 8...\}$ is, $P(W = w_i) = \sum_{j=\lfloor \log(w_i)+1 \rfloor}^{\infty} \frac{1}{j} \left(F(2^j - 1) - F(2^{j-1} - 1) \right)$ where $F(j)$ is the distribution function of the file size in terms of packets. The expression $\lfloor \log(w_i) + 1 \rfloor$ corresponds to the number of rounds which are needed in order to reach the congestion window w_i. Because of the monotonicity of the congestion window evolution during slow start, the probability of having window size w_i given the session lasts j rounds is simply equal to $1/j$. Next assuming the batches arrive according to a Poisson process and the service time is exponentially distributed we form the transition matrix of the corresponding Markov process and compute the steady state distribution π_i, $i = 0, ..., K$, where K is the bottleneck buffer size. Then we compute the packet loss probability p as follows:

$$p = \frac{1}{E[W]} \sum_{i=0}^{K} \pi_i \sum_{w_j = K-i+1}^{\infty} (w_j - K + i) P(W = w_j).$$

In Figure 21 we plot the above packet loss probability for the batch model as a function of load in the case of short TCP flows. In the same Figure we also plot the packet loss probability given by the $M/M/1/K$ model (3).

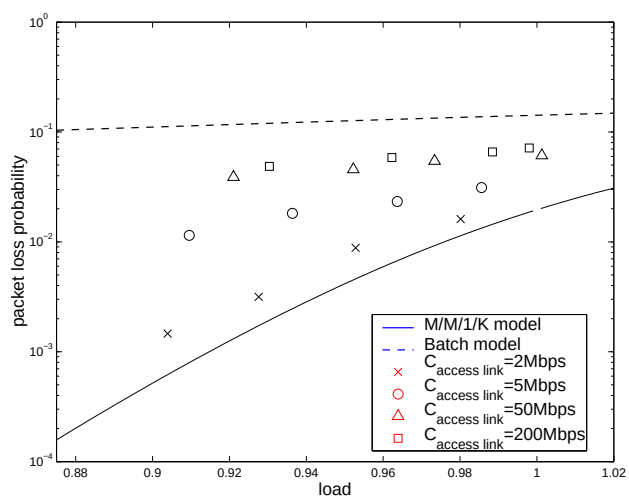

Figure 21: Bounds on packet loss probabilities

All points obtained from NS simulations correspond to the following set of nominal loads $\rho_0 = \{0.9,\ 0.925,\ 0.95,\ 0.975\}$. The $M/M/1/K$ and the Batch models provide indications for lower and upper bounds for the packet loss probability, respectively. We observe that for a value of NC_{ac}/C around 2 (i.e. packet arrival process close to Poisson) the loss probability is close to the lower bound provided by the $M/M/1/K$ model.

5 Discussion

We analyze in this paper the performance of a bottleneck link crossed by TCP traffic. We give a particular attention to the packet arrival process at the input of the bottleneck link. The campaign of simulations we run showed that this process is close to Poisson when the number of access links is large and when the speed of these links is slow. This observation scales well with the bandwidth of the bottleneck and is insensitive to the distribution of file sizes. For a certain number of access links, we find that there is some speed of access links that makes the packet arrival process very close to Poisson. We give an empirical expression for this threshold speed.

Motivated by the Poisson property of packet arrivals, we propose a model for the bottleneck link using a Fixed Point Approach. The buffer in the bottleneck router is modeled by a M/M/1/K queue and the TCP traffic is modeled using only its average rate. We introduce a generic function that models the throughput of long-lived TCP connections when the average queue size and the packet loss rate in the bottleneck router are given. Our simulations show that this model for the network provides good results when the packet arrival process is close to Poisson. When we are far from Poisson, the traffic is bursty and the model underestimates the packet loss rate in the bottleneck. For this purpose, we propose an upper bound on the packet loss rate, obtained by using a batch model for the packet arrival process.

One interesting finding of our work is the limitation of NS when we simulate fast speed links. To our knowledge, nobody has observed this problem before. The granularity of the time in NS is the microsecond, so the accuracy of the simulator is questionable when many events are separated by less than one microsecond. For packets of 500 bytes size, the accuracy of the simulator is bad when 1 Gbps links are considered. The smaller the packet size, the slower the maximum link speed we can use. We believe that the time granularity of NS needs to be improved if we want this simulator to be used as a reference in simulating network protocols in high speed environments.

Our work can be extended in different directions. One direction is to consider other buffer management policies than Drop Tail, for example RED. Another direction is to consider real distributions for packet sizes, for example deterministic or multimodal. Also more research is needed to understand the interaction between short and persistent TCP flows. Some UDP traffic can also be considered in addition to the TCP traffic. The presence of TCP traffic may improve the multiplexing of packets in the bottleneck router. The result of the model in terms of packet loss rate and average queue size can be used for network dimensioning issues. One can plug the result in a expression for TCP latency and dimension the bottleneck router so that the average latency is less than a certain value. The dimensioning can also be done so as to guarantee some Quality of Service for multimedia applications.

References

[1] E. Altman, K.E. Avrachenkov and C. Barakat, "A stochastic model of TCP/IP with stationary random losses", *ACM SIGCOMM 2000*, Stockholm, Sweden, also in *Computer Communication*

Review, v.30, no.4, pp.231-242, 2000.

[2] C. Barakat, P. Thiran, G. Iannaccone, C. Diot and P. Owezarski, "A flow-based model for Internet backbone traffic", Technical Report EPFL/DSC/01/44, Jul. 2001. Also a shorter version appears in *ACM SIGMETRICS 2002*.

[3] J-C. Bolot, "End-to-end packet delay and loss behavior in the Internet", *Proc. ACM Sigcomm '93, pp. 289-298*, San Francisco, CA, Sept. 1993.

[4] T. Bu and D. Towsley, "Fixed point approximation for TCP behaviour in an AQM network", in Proceedings of SIGMETRICS'2001 conference.

[5] The web site of the Cooperative Association for Internet Data Analysis (CAIDA): http://www.caida.org/.

[6] C. Casetti, M. Meo, "A New Approach to Model the Stationary Behavior of TCP Connections", *IEEE INFOCOM*, Tel Aviv, Israel, March 2000.

[7] Jin Cao, William S. Cleveland, Dong Lin, Don X. Sun, "On the nonstationarity of Internet traffic", *ACM SIGMETRICS*, June 2001.

[8] S. Dawkins, G. Montenegro, M. Kojo, V. Magret, N. Vaidya, "End-to-end Performance Implications of Links with Errors", August 2001.

[9] S. Floyd, S. Ratnasamy, and S. Shenker, "Modifying TCP's Congestion Control for High Speeds", http://www.icir.org/floyd/hstcp.html.

[10] M. Garetto, R. Lo Cigno, M. Meo, M. Ajmone Marsan, "A Detailed and Accurate Closed QUeueing Network Model of Many Interacting TCP Flows," IEEE Infocom 2001, Anchorage, Alaska, USA, April 22-26, 2001.

[11] R.J. Gibbens, S.K. Sargood, C. Van Eijl, F.P. Kelly, H. Azmoodeh, R.N. Macfadyen, N.W. Macfadyen, "Fixed-point models for the end-to-end performance analysis of IP networks," 13th ITC Specialist Seminar: IP Traffic Measurement, Modeling and Management, Sept 2000, Monterey, California.

[12] M. Mathis, J. Semke, J. Mahdavi, and T. Ott, "The Macroscopic Behavior of the TCP Congestion Avoidance Algorithm", *ACM Computer Communication Review*, Jul 1997.

[13] R. Nelson, "Probability, Stochastic Processes, and Queueing Theory", Springer-Verlag, 1995.

[14] The web site of the National Laboratory for Applied Network Research (NLANR): http://www.nlanr.net/.

[15] http://www.isi.edu/nsnam/ns/index.html. Release 2.18a simulations.

[16] J. Padhye, V. Firoiu, D. Towsley, and J. Kurose, "Modeling TCP throughput: A simple model and its empirical validation", *ACM SIGCOMM*, Sep 1998.